Multi network classification scheme for detection of colonic polyps in CT colonography data sets

Anna K. Jerebko*, James D. Malley, Marek Franaszek, Ronald M. Summers
National Institutes of Health, Bethesda

ABSTRACT

A multi-network decision classification scheme for colonic polyp detection is presented. The approach is based on the results of voting over several neural networks using different variable sets of size N which are selected randomly or by an expert from a general variable set of size M. Detection of colonic polyps is complicated by a large variety of polypoid looking shapes (haustral folds, leftover stool) on the colon surface. Using various shape and curvature characteristics, intensity, size measurements and texture features to distinguish real polyps from false positives leads to an intricate classification problem. We used 17 features including region density, Gaussian and average curvature and sphericity, lesion size, colon wall thickness, and their means and standard deviations in the vicinity of the prospective polyp. Selection of the most important parameters to reduce a feature set to acceptable size is a generally unsolved problem. The method suggested in this paper uses a collection of subsets of variables. These sets of variables are weighted by their effectiveness. The effectiveness cost function is calculated on the basis of the training and test sample misclassification rates obtained by the training neural net with the given variable set. The final decision is based on the majority vote across the networks generated using the variable subsets, and takes into account the weighted votes of all nets. This method reduces the false positive rate by a factor of 1.7 compared to single net decisions. The overall sensitivity and specificity rates reached are 100% and 95% correspondingly. Best specificity and sensitivity rates were reached using back propagation neural nets with one hidden layer trained with the Levenberg-Marquardt algorithm. Ten-fold cross-validation is used to better estimate the true error rates.

Keywords: virtual colonoscopy, colon cancer, classification, neural network

1. INTRODUCTION

Computed tomographic colonography (CTC) as an alternative colon cancer screening has progressed rapidly over the past 6 years [1]. Colon cancer remains a danger for approximately 6% of Americans who may face this problem during their lifetime. Total colon evaluation by widely accepted methods like barium enema and conventional colonoscopy can identify most of the polyps before they progress to cancer. Nevertheless a large portion of the population over 50 years of age does not undergo a screening process because of the fear and discomfort of the screening test. The relatively high cost of the exam and difficulties in reaching the most distant parts of the colon are a problem with the gold standard test, conventional colonoscopy. For these reasons, there is a need to develop modern, less invasive, inexpensive and potentially more powerful screening methods, such as CTC. Computer-aided detection (CAD) is important in CTC interpretation and diagnosis, since the interpretation times are still 15 to 20 minutes per patient, which increases the cost of the examination.

The classification method to detect colonic polyps suggested in this paper combines two techniques: a shape-based primary classifier and a higher-level classifier based on an aggregate of back propagation neural networks [2]. Three-dimensional shape has a key role in human visual perception of an object in general and in colonic polyp detection specifically. For example, high curvature and roundness appear to be indispensable conditions for polyp detection of both conventional colonoscopy and CTC. The shape-based classifier for primary lesion detection allows one to eliminate more than 95 % of the colonic surface from consideration [5]. The remaining 5% represents on average 300 candidate lesions per colon, which are a mixture of real polyps (approaching 100% sensitivity) and false positive detections. These remaining candidate polyps are suitable for higher-level classification.

* anya@nih.gov; phone 301 496 7700; fax 301 496 9933; Dept of Radiology, NIH, Bethesda, MD 20892-1182

Whereas using the shape criteria as a primary feature seems to be obvious, other features taken into consideration by radiologists in the polyp diagnosis process seem to be more difficult to quantify. In this situation it is profitable to use classifiers that can learn from the data and use as many features as possible to make a decision. We propose here a classification scheme containing the aggregate of several neural networks (NNs), also called a "NN forest", using different sets of 4 features each selected from the general list of up to 17 features. The decision in such a composite classifier is based on the majority votes of its component NNs. The advantage of this approach is that none of the NNs are overloaded with the input data and at the same time the classifier takes into consideration enough distinguishing features to provide sufficiently high sensitivity and specificity rates.

The rest of the paper is organized as follows. Our classification algorithm is introduced in Section 2. In Section 3 experimental data is described. The performance analysis is presented in Section 4. Section 5 contains the discussion.

2. METHODS

1.1 Shape classification

The first step of the classification algorithm is meant to eliminate the majority of the colonic surface (up to 95%), which is unlikely to contain polyps. It can be done by analyzing the shape features of the colonic surface such as curvature and sphericity. Although colonic polyps may vary in size and shape, e.g. pedunculated, hyperplastic and sessile ones [3], the majority of them appear as bumps on the rendered colon surface.

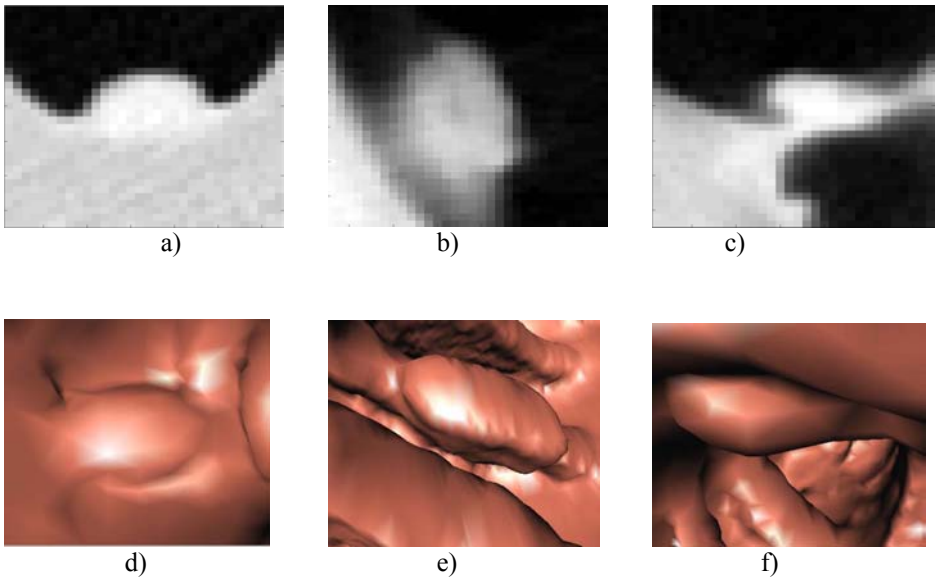

Fig. 1: (a) Sessile polyp, (b) Polyp on a fold, (c) Pedunculated polyp lying on its side, (d), (c), (f) Three-dimensional images of the same polyps

The algorithm first selects lesions having elliptical curvature and sphericity above certain thresholds. Then other important characteristics are calculated, such as size, region density, wall thickness, etc., which make possible improved specificity.

1.2 Aggregate of back propagation NNs

The next step in the classifier is based on various texture, density, geometric parameters of the lesions, colon surface and wall characteristics, their means and standard deviations. The parameters listed above make a list of more than 17 features useful for classification. On the one hand, taking more features into consideration allows more precise discrimination between the real polyps and the false positives detections. On the other hand using too many features unacceptably increases the complexity of the model because the number of hidden neurons of neural net classifier

corresponds to the dimensionality of the feature space. Keeping the dimensionality of the feature space small effectively controls the model complexity [6]. Here we suggest breaking the set of features into subsets and using the combination of several simple classifiers each processing a small number of input features. This approach combines the advantages of using the large number of features and keeping the feature space small for each neural net in the forest. The entire polyp detection scheme is depicted in Figure 2.

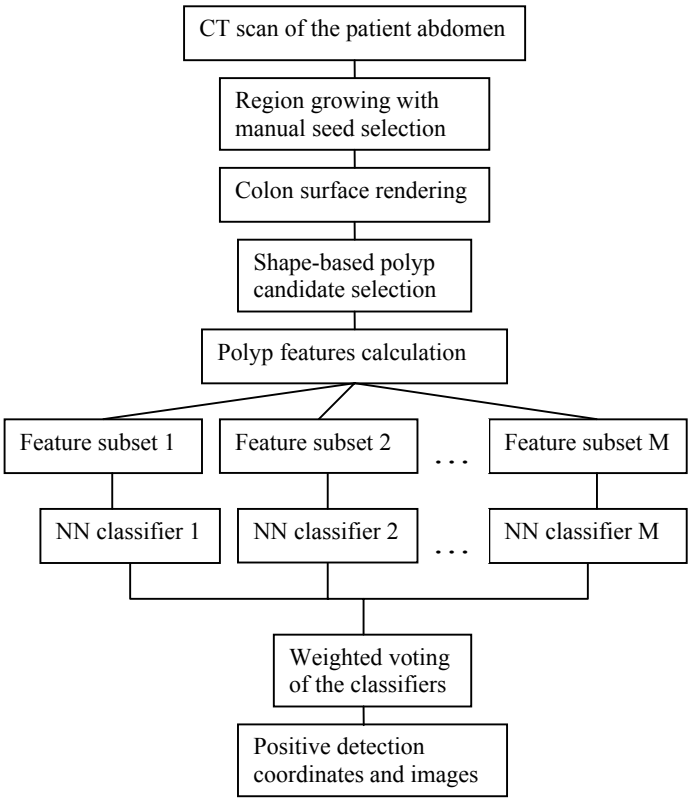

Fig. 2: Polyp detection scheme

The feature space F is divided into M subsets, each subset contains N features:

$$F \Rightarrow \begin{bmatrix} f_{11} \\ f_{12} \\ \dots \\ f_{1N} \end{bmatrix}, \begin{bmatrix} f_{21} \\ f_{22} \\ \dots \\ f_{2N} \end{bmatrix} \dots \begin{bmatrix} f_{M1} \\ f_{M2} \\ \dots \\ f_{MN} \end{bmatrix}; \quad (1)$$

These subsets are the input for the component classifiers NNs. The features may be repeated in the different subsets, so as many combinations of them as possible are used. The NN classifiers represent multilayer perceptrons trained with a backpropagation algorithm. Each NN consists of an input layer with N input neurons, an output layer with 2^Z hidden neurons and an output neuron. The hidden neurons are connected to all input neurons and an output neuron. The polyp data set contains *K* samples. *M* feature vectors of size *N* are calculated for each polyp. The output O_{jk} of the *j-th* hidden neuron when *k-th* sample *X* is presented is calculated as follows:

$$O_{jk} = \sigma(\sum_{i=1}^{N} w_{ij} * X_{ik} + b_j); \quad (2)$$

Where w_{ij} is the weight of connection, b_j is the bias of connection, j=1…2^Z and σ is a sigmoid transfer function.

The output of the network g_k when *k-th* sample is presented is calculated as follows:

$$g_k = \sum_{j=1}^{2^Z} w_j^{out} * O_{jk} + b_{out}; \quad (3)$$

where w_j^{out}, b_{out} are the weights and the bias of the output neuron connections to all neurons of the hidden layer.

The goal of the training process is to have the NN output equal -1 when the sample presented to the input layer represents a false positive detection and 1 when a true positive detection is presented.

Let *V* denote the weights and the biases of the NN: $V(w_{ij}, b_j, w_j^{out}, b_{out})$.

The risk functional is the mean-squared error (MSE) function:

$$R = \frac{\sum_{k=1}^{K} (g(X_k, V) - y_k)^2}{K}; \quad (4)$$

where $X_k(f)$, y_k is the *k-th* sample, $y_k = 1$ if $X_k(f)$ corresponds to a true positive, and $y_k = -1$ if $X_k(f)$ corresponds to a false detection.

We use the Nguyen-Widrow method [4] to initialize *V*. An advanced non-linear Levenberg-Marquardt optimization algorithm [5] is used to train weights and biases *V*. Training is an iterative process and stops when the desired mean square error is reached.

1.3 Voting system for NN forest

The classification model containing trained NNs produces the classification decision on the basis of weighted voting of all NNs in such a way that the NNs with worst performance make less contribution to the final decision. The weight of each NN is calculated on the basis of the numbers of false positive and false negative responses found upon applying the trained NN to training and test sets. The weights (*P*) are calculated using:

$$P = \frac{1}{\frac{k_1 * Fn_{tr}}{N_1} + \frac{k_2 * Fn_{test}}{N_2} + \frac{k_3 * Fp_{tr}}{N_3} + \frac{k_4 * Fp_{test}}{N_4}}; \quad (5)$$

where k_i , i=1,..4, are coefficients adjusted according to clinical needs. For better sensitivity k_1 and k_2 should be higher than k_3 and k_4 , and if specificity is more important for the particular study, then k_3 and k_4 should be higher. N_1 , N_2 , are the numbers of polyps in training and test set. N_3 , N_4 are the numbers of nonpolyps in training and test set, respectively. Fn_{tr} , Fn_{test} are the numbers false negative responses found upon applying the trained NN to training (Fn_{tr}) and test sets (Fn_{test}). Fp_{tr} , Fp_{test} are the numbers of false positive responses in the results of applying the trained NN to training (Fp_{tr}) and test sets (Fp_{test}).

To get the best estimates of NN's weights we used a ten-fold cross-validation scheme: each NN was trained ten times with the use of 90% of polyps and nonpolyps samples as a training set and remaining 10 percent of them as a test set in such a way that each polyp and nonpolyp was presented once in the test set. The final NN's weights were obtained by averaging over all 10 splits. The NN forest classification decision *D* is calculated as follows:

$$D = \sum_{i=1}^{M} g_i * P_i; \quad (6)$$

where g_i is the output of the *i-th* NN, P_i is the *i-th* NN's weight, and M is the number of NNs in the forest.

3. EXPERIMENTAL DATA

The polyp database used for training and test purposes in our experiment was obtained from 80 studies containing supine and prone screening of 40 patients. CT scans were done on G.E.Lightspeed scanners. Scanning parameters were 120kVp, 50mAs (mean), field of view to fit (38-46 cm), 5 mm collimation, HQ mode, and 3 mm reconstruction interval (2mm overlap). We used colonoscopic examination of the same patients matched to the CT colonography as a ground truth to obtain the coordinates of the true positive detections. 39 polyps from 0.3 to 2.5 cm in size a well as 25012 polypoid-looking sites were selected by our endoscopic research software [6] using the shape criteria.

4. PERFORMANCE ANALYSIS

We used the same cross-validation scheme as described in Section 2C to analyze the sensitivity, specificity and stability of NN forest and single NNs. Cross-validation for each NN and the forest was done at 13 different mean square error (MSE) levels from 10^{0} to 10^{-3} . To reduce the computational cost of the analysis, the forest consisted of four NNs having the highest weights. Nevertheless, calculations took 72 hours on an SGI Origin 3400 supercomputer with eight MIPS R12000 400 MHz CPU's.

Analysis has shown that at high MSE levels when $10^{0} \leq MSE < 10^{-1}$, NN's and forests classify the majority of samples as nonpolyps, giving high specificity and low sensitivity. As can be seen from Figure 2a, the forest specificity in the interval $10^{0} \leq MSE < 10^{-2.25}$ is higher than the average specificity of a single NN. The forest sensitivity is higher than that of an average single NN in the interval $10^{-0.75} \leq MSE < 10^{-2.75}$. Sensitivity reaches its maximum for single NNs and NN forest at MSE level $10^{-1.75}$. At this point the forest sensitivity averaged over 10 cross-validation steps is 100% with 0 standard deviation (STD), while average single NN gives 95% sensitivity with 11.5% STD (Table 2). Specificity of forest at this point is 1% higher than the one on single NN, that means 17.7% improvement in number of false positives per study. Both NN forest and single NN classification models become over trained with MSE levels lower than $10^{-1.75}$ and sensitivity rates decrease rapidly. This is the well-known problem of overfitting, where the classification model works perfectly on the training set and gives poor results on the test set. The fact that the STD of both specificity and sensitivity of the forest stays higher than those of the average NN at all MSE levels up to $10^{-2.25}$ (where NNs become over-trained) testifies to the better stability of the forest over the single NN.

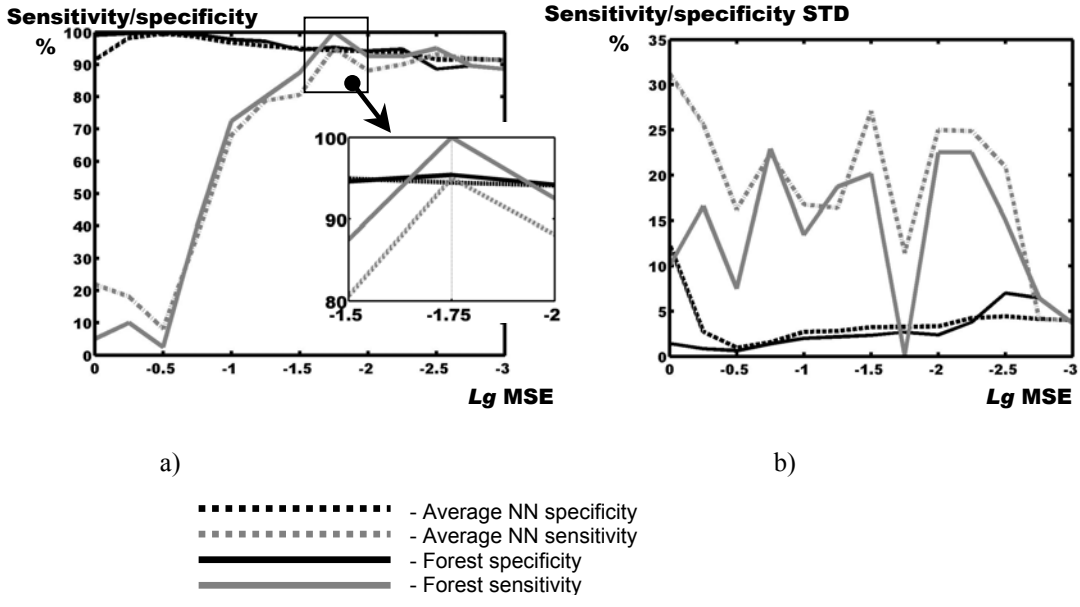

Fig. 3: (a) Cross-validated forest and average single NN specificity and sensitivities
(b) Standard deviation of forest and average single NN specificity and sensitivities

Table 2: Performance of the NN forest and the average single NN at MSE level $10^{-1.75}$

	Number of false negatives (overall)	Number of false positives (per study)	Specificity	Specificity STD	Sensitivity	Sensitivity STD
NN Forest	0	14	95.4%	2.7%	100%	0
Average single NN	0.2	17	94.4%	3.2%	95%	11.5%

5. DISCUSSION

Although better specificity rate (98.5%, 3-5 false positives per study compared to 17 here) with single NN could be reached with a larger training set containing 2000 nonpolyp samples and 39 polyps, in this study for evaluation of benefit of using NN aggregate over the single NN, the smaller training set of 500 nonpolyps and 39 polyps was used to reduce the training time. Intensive validation schemes providing the best estimates of error rates are usually computationally expensive. In this study 520 NNs were trained with different levels of mean square error (MSE) (from 10^{0} to 10^{-3}). Training one NN to reach a higher level of MSE than 10^{-2} usually takes less than 25 minutes on Pentium III, which makes it suitable for clinical purposes. When higher specificity becomes more important, it is recommended to use bigger data sets for training. The investigation described in this paper confirms that using an aggregate of classifiers such as NNs may improve classification results compared to using a single classifier.

ACKNOWLEDGMENTS

We thank C. Daniel Johnson, MD of the Mayo Clinic for providing CT colonography data sets.

REFERENCES

1. Summers, R.M., *et al.*, "CT and MR Colonography: Summary of Progress from 1995 to 2000. Current Problems in Diagnostic Radiology", 2001: p. 145-167
2. O'Brien, M.J., *et al.*, "The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas". Gastroenterology, 1990. **98**(2): p. 371-9.
3. Nguyen, D., and B.Widrow. "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights". in *International Joint Conference on Neural Networks*. 1990.
4. More, J.J., "The Levenberg-Marquardt algorithm: implementation and theory". *Numerical Analysis, Lecture Notes in Mathematics 630,*, ed. G.A. Watson. 1977, Heidelberg: Springer-Verlag. 105-116.
5. Summers, R.M., *et al.*, "Automated polyp detector for CT colonography: feasibility study". *Radiology*, 2000. **216**(1): p. 284-90.
6. Cherkassky, V.S., Mulier F., *Learning from data. Concepts, Theory, and Methods*. A Wiley-Interscience publication, 1998.